

AperTO - Archivio Istituzionale Open Access dell'Università di Torino

Constituent order in compounds and syntax: typology and diachrony

This is the author's manuscript

Original Citation:

Availability:

This version is available <http://hdl.handle.net/2318/130960> since

Published version:

DOI:10.1007/s11525-009-9125-x

Terms of use:

Open Access

Anyone can freely access the full text of works made available as "Open Access". Works made available under a Creative Commons license can be used according to the terms and conditions of said license. Use of all other works requires consent of the right holder (author or publisher) if not exempted from copyright protection by the applicable law.

(Article begins on next page)

Constituent order in compounds and syntax: typology and diachrony

Livio Gaeta

Received: 7 April 2008 / Accepted: 3 April 2009 / Published online: 5 May 2009
© Springer Science+Business Media B.V. 2009

Abstract The paper addresses the question of the correspondence between constituent order in compounds and in syntax. While a strictly synchronic perspective does not lead us to any significant generalization as ascertained by Bauer (Language typology and language universals, Mouton de Gruyter, Berlin 2001), adopting a diachronic point of view allows us to formulate the question in general terms by making reference to the logical problem of what is the transition permitted from a certain synchronic stage to another. On the basis of a large language sample, it is shown that constituent order in compounds heavily relies on syntax. This must be understood in the terms of a diachronic universal reflecting Hawkins' (Word order universals, Academic Press, New York 1983) Double Acquisition Hypothesis. For this specific property of compounds, morphology does not seem to be autonomous from syntax, albeit the relation between morphology and syntax must be thought of as a multi-faceted one.

Keywords Compounds · Constituent linearization · Syntax

L. Gaeta(✉)
Università di Napoli, Napoli, Italy
E-mail: livio.gaeta@unina.it

“Die eigentlich normale Entstehungsweise alles Formellen in der Sprache bleibt daher immer ... die Komposition”¹ (Paul [1920] 1995, p. 325).

1 Introduction

That there is a strict relation between the constituent order in compounds and in syntax is an old hypothesis, which goes back at least to the second half of the nineteenth century, as summarized in diachronic or—better—genetic terms by Paul (1920, pp. 5–6). To his mind, compounds develop out of syntactic units via a mechanism of univertation, as illustrated by German compounds like *Tageslicht* ‘daylight’ or *Sonntag* ‘sunday’ which go back to the Old High German compounds *tageslieht* and *sunnūntag*, and respectively result from the noun phrases *tages lioht* ‘day’s light’ and *sunnūn tag* ‘sun’s day’. The univertation is taken to be a consequence of the isolation of the units with respect to their constituting elements, which may be due to semantic opacity (for instance, the meaning of the German compound *Nasenbein* ‘nasal bone’ is obscured by the semantic change of Old High German *bein* ‘bone’, whose modern correspondent only means ‘leg’), and/or to morphological (for instance, the ending *-n* in *Nasenbein* used to be a genitive marker) or phonological reasons (for instance, the ending *-es* in *Tageslicht* cannot be shortened as in the genitive phrase *des heutigen Tags* ‘of the present day’).

Paul postulates the intervention of an analogical mechanism of pattern generalization on the basis of a handful of frozen units, which leads to their reanalysis in terms of a pattern for compounding. Notice that he extends his analysis also to those cases for which no historical evidence is available, for instance the Old High German stem compounds like *tag-a-stern* ‘morning star’ or *pir-o-baum* ‘pear-tree’: “Wenn wir aber auch die Entstehung der älteren Zusammensetzungen nicht verfolgen können, so unterliegt es doch keinem Zweifel, daß dieselbe sich in der gleichen Weise vollzogen hat wie die jüngeren” (Paul 1920, p. 6).² Given that compounds genetically are the result of frozen syntactic units, they are likely to reflect the constituent order of their primary sources, as shown by the examples reported above which display the normal Old High German genitive/noun order, subsequently changed into the modern noun/genitive order (cf. Demske 2001, pp. 296–316 for a recent discussion of the question).

Since then, the idea has been circulating that the linearization order of constituents in compounds reflects the order of constituents in syntax, albeit the synchronic relationship may be obscured by subsequent syntactic change.

On the other hand, theoretical morphologists have approached the question from a strictly synchronic perspective trying to assess the question in purely structural terms. From this perspective, the interest lies in possible universal constraints on the

¹ ‘The actually normal way whereby all formal structure arises in the language turns out to be always ... compounding’.

² ‘Even if we cannot pursue the origin of the older compounds, there is no doubt that it took place along the same lines of the newer ones’.

architecture of grammar, which are held to be responsible for the linearization of constituents in compounds as well as in syntax.

In this paper, both viewpoints will be reviewed, with the aim of verifying the basic tenet of an alleged correspondence between constituent order in compounds and syntax with the help of a sample of 50 languages. We will first start in Sect. 2 with discussing the synchronic perspective, then we will move in Sect. 3 to the diachronic point of view, trying to revive the old correspondence hypothesis in more modern terms, which benefit from the Greenbergian typological turn. Then, we will discuss in Sect. 4 the basic methodological issues underlying the language sample which has been collected in order to validate the correspondence hypothesis. The results of the investigation are discussed in Sect. 5 and seem to strongly support the idea of a correspondence between syntax and morphology. This has to be understood in terms of a diachronic universal, very much in the sense of Hawkins' (1983) Double Acquisition Hypothesis, as argued in Sect. 6. The final Sect. 7 suggests possible explanations for the survival of the cases of lacking correspondence in spite of their apparently dysfunctional status.

2 The synchronic point of view

Is there any relation between constituent order in compounds and in syntax? As far as a strictly synchronic viewpoint is concerned, there are basically two ways of approaching the question. The first one is to look at the empirical evidence on the basis of a wide language sample. As odd as this may appear, there is not much research in this direction. One exception is Bauer (2001), who investigates compounds on a large varieties of language.

In spite of the optimistic stance of the expectations (cf. Bauer 2001, p 697: “[i]t might be expected that the order of head and modifier nouns in a nominal compound should reflect the order of noun and adjective”), the conclusion is rather disappointing, because “this is not necessarily the case” (Bauer 2001, p. 697). Nonetheless, “it seems that there is a slight preference for modifier noun + head structures, independent of the syntactic order of adjective and noun” (Bauer 2001, p. 697). A similar conclusion is reached when the constituent order in compounds is compared to a different syntactic type, namely the genitive/noun constructions: “The order of modifying and head element in a compound most often seems to reflect the order of possessor and possessed in noun phrases, but the order of modifier and head is frequently variable in compounds” (Bauer 2001, p. 705). He finds out some evidence pointing to a negative answer of the question of the dependence of constituent order in compounds on syntax. Namely, a preference for right-headedness is observed, which is independent of the linearization in syntax and emphasizes the autonomy of compounds as morphological objects.

Furthermore, if compounds with a verbal element and a nominal argument, the so-called synthetic compounds, are considered, the thesis of the autonomy of compounds from syntax is strengthened. Again, the hypotheses basically test whether constituent order in compounds reflects syntactic principles:

There are at least two hypotheses which might be put forward to explain the ordering of verbal element and nominal argument in a synthetic compound. (a) It might be that the order of verb and argument corresponds to the order of verb and direct object in the syntax ... (b) It might be that the order of verb and argument corresponds to the order of head and modifier in tatpuruṣa [i.e., determinative] compounds in the same language. (Bauer 2001, p. 701)

The conclusion is that “hypothesis (b) is more strongly supported than hypothesis (a), ... the ordering in a synthetic compound relies more on morphological principles than on syntactic ones” (Bauer 2001, p. 702).

A second way of approaching the question is of a rather speculative character. Namely, universal hypotheses on the architecture of grammar are formulated which account for language-specific properties. This is typical for theoretical approaches to universals of a rather formal character, which are in many cases stipulative in the sense of being strictly connected to their underlying theoretical framework (cf. Gaeta 2005 for a critical survey). As for constituent order in compounds, the so-called Right-Hand Head Rule has been suggested by Williams (1981), who explicitly assumes the right position of the head to be universal (cf. Spencer 1991, pp. 187–190 for a discussion).³ Notice that Dressler (2005, p. 32) reaches a similar conclusion on empirical grounds, at least for a specific subgroup of compounds, namely the determinative type:

All of the above compounds [i.e., N+N determinative compounds] have their head on the right, which corresponds to a universal preference, called the right-hand head rule by Williams.

Unfortunately, we are not precisely told what the empirical basis is for postulating such a universal preference.

To draw a provisional conclusion on the basis of the existing literature, we may say that the relation between constituent order in syntax and in compounds is unclear. Furthermore, there appears to be a tendency for morphology to win over syntax: linearization in compounds is claimed to obey morphological rather than syntactic principles. A universal preference for right-headed compounds is assumed, which is independent of any syntactic constraint.

3 A diachronic approach to constituent order in compounds

Let us come back to the traditional hypothesis mentioned above, which assumes a consistent linearization pattern for compounds and phrases, because the former genetically rely on the latter via a mechanism of univerbation. We can try to give a synchronic sense to this diachronic hypothesis by taking into consideration the contribution of Greenbergian typology. An important consequence of language

³ Spencer (1991, p. 188) points out that “[t]he Righthand Head Rule itself embodies an extremely strong universal claim about word structure, which on the face of it is simply wrong”. In Williams’ approach all inflectional affixes are assigned a head role, and therefore the hypothesis predicts that there should be no languages in which inflections can be prefixes, which is utterly false.

typology is its impact on depicting possible scenarios of language change. In this regard, starting with a general implicational universal of the type “If a language has property P, then it has property Q”, Hawkins (1990) draws the logical conclusion that such a universal implies a strong restriction on the diachronic variation allowed to a natural language:

- (1) a. $P \ \& \ Q \ > \neg P \ \& \neg Q$
 b. $P \ \& \ Q \ > \neg P \ \& \ Q \ > \neg P \ \& \neg Q$
 c. $P \ \& \ Q \ > *P \ \& \neg Q \ > \neg P \ \& \neg Q$

For a language displaying both properties P and Q, only two possible diachronic scenarios are available: either this language may change both properties simultaneously as in (1a), or P may be changed before Q as in (1b). It should never be possible that a diachronic scenario involves a stage in which Q is changed before P, as in (1c). To quote a real case from word order, the universal suggested by Hawkins (1983, p. 67), according to which “[i]f a language has Postp[ositional] word order, and if the adjective precedes the noun [i.e., P, LG], then the genitive precedes the noun [i.e., Q, LG]”, implies the following restriction on possible diachronic scenarios:

- (2) a. $[\text{Adj N}]_{\text{NP}} \ \& \ [\text{Gen N}]_{\text{NP}} \ > [\text{N Adj}]_{\text{NP}} \ \& \ [\text{N Gen}]_{\text{NP}}$
 b. $[\text{Adj N}]_{\text{NP}} \ \& \ [\text{Gen N}]_{\text{NP}} \ > [\text{N Adj}]_{\text{NP}} \ \& \ [\text{Gen N}]_{\text{NP}} \ > [\text{N Adj}]_{\text{NP}} \ \& \ [\text{N Gen}]_{\text{NP}}$
 c. $[\text{Adj N}]_{\text{NP}} \ \& \ [\text{Gen N}]_{\text{NP}} \ > *[\text{Adj N}]_{\text{NP}} \ \& \ [\text{N Gen}]_{\text{NP}} \ > [\text{N Adj}]_{\text{NP}} \ \& \ [\text{N Gen}]_{\text{NP}}$

In other words, it should never be possible that a diachronic scenario involves a stage like (2c) in which the word order $[\text{Adj N}]_{\text{NP}}$, namely P, cooccurs with the order $[\text{N Gen}]_{\text{NP}}$, namely $\neg Q$, in a language displaying postpositions, while it is possible that $\neg P$, namely $[\text{N Adj}]_{\text{NP}}$, cooccurs with Q, namely $[\text{Gen N}]_{\text{NP}}$ as in (2b). This diachronic prediction is sustained by the synchronic observation that no language displays the word order correlation $*[\text{Adj N}]_{\text{NP}} \ \& \ [\text{N Gen}]_{\text{NP}}$ (cf. Hawkins 1983, p. 67 for the discussion of two apparent counterexamples).

Notice that these diachronic scenarios outline developmental schemes, which exclude a certain configuration as a possible evolutionary stage. To give an empirical content to his approach, Hawkins formulates the Double Acquisition Hypothesis, whereby given the impossible cluster of properties $*P \ \& \neg Q$, P cannot be acquired in the total absence of Q, and requires an evolutionary stage in which either Q is already present at the earlier stage (for instance as a doublet of $\neg Q$) or is acquired as doubling structure simultaneously with P.⁴

⁴ Cf. Hawkins (1983, p. 213): “Given a set of synchronic universal implications of the form ‘if P then Q’, where P and Q are basic word orders of certain specified types; then, at two successive stages in the growth of a language,

If: P is acquired as a doubling structure from the earlier uniquely $\neg P$ stage;

Then: Either Q must already be present at the earlier stage (whether as a doublet with $\neg Q$ or not), or, if it is not present, Q must be acquired as a doubling structure simultaneously with P. But P will not be acquired in the total absence of Q”.

This hypothesis suggests that certain configurations are impossible in given developmental schemes, in other words as transitional stages. This observation will turn out to be useful when we will try to understand why unexpected mismatches occur.

After this premise, let us now discuss the traditional issue of the correspondence hypothesis. A convenient starting point is provided in the usual elegant way by Comrie (1980, p. 85):

Another area where one might expect the order of morphemes to represent, in a frozen state, the earlier order of words is with compounds, such as English *blackbird*, where the order Adjective-Noun reflects the earlier, and still current, order of words ... Although the English Adjective-Noun compounds reflect equally the synchronic word order, this is not true of compounds with a verb or verbal derivative as head and an object noun phrase or adverbial as adjunct: such compounds have the verb in final position ... *nutcracker* ... Given the observation that English has developed from a basic language with basically verb-final word order (SOV) to a language with basic SVO order, we can say that the current order in these compounds does indeed reflect the earlier verb-final word order.

This long quote contains both the hypothesis and its falsification, namely the existence of non-correspondent structures. However, the latter are supposed to be straightforwardly explained in diachronic terms by postulating a stage like (1b) above, namely an antecedent stage in which the modifier/head order in the compound necessarily cooccurred with the OV order. OV order was subsequently changed giving rise to the synchronic mismatch. In other words, on the basis of the logical implications sketched in (1) above, the universal diachronic hypothesis is derived which can be formulated adopting Harris and Campbell's (1995, p. 203) terms:

Hypothesis 1 In Language L, the order of words in compounds that are coined at time t_i is the same as the order of words used in phrases at some time before t_i .

On this hypothesis Harris and Campbell (1995, p. 203) comment quite pessimistically that

It is possible that hypothesis [1] is correct, but it is an untestable hypothesis. There is no data that could, even hypothetically, disprove it. Even if hypothesis [1] could be shown to be true, it still could not serve as a tool of reconstruction, since we would not know how far back we were reconstructing, and could not relate a dyadic order reconstructed in this way to other features of the language.

There are actually two different observations implicitly contained in this comment. The first one is of a methodological nature, and reflects the attitude espoused by historical linguistics since the nineteenth century, namely the comparative-reconstructive view. In this view, even admitting that the hypothesis holds true, it does not explicitly say when this hypothetical antecedent stage can be reasonably assumed to occur. It might be that the antecedent stage occurred immediately before

the occurring state of affairs, or much earlier.⁵ In other words, from a reconstructive point of view, the correlation between constituent order in compounds and syntax does not help much for establishing the internal chronology of changes to be assumed for a certain language, because it does not correlate with other independent features. For this unreliability in terms of internal reconstruction, historical linguistics discarded much of the empirical value contained in this reconstructive method, as Harris and Campbell correctly maintain. More generally, doubts can be raised against the feasibility of directly applying the comparative-reconstructive method to syntax (cf. Harrison 2003, pp. 225–227 for a critical discussion).

However, this negative attitude against the reconstructive usefulness of the method should not be extended to the second observation contained in Harris and Campbell's comment, namely that Hypothesis 1 is supposed to be empirically untestable. As a counterthesis, we can formulate the following Hypothesis 2:

Hypothesis 2 In Language L, the order of words in compounds that are coined at time t_i is likely to be completely independent from the order of words used in phrases at time t_i or at some time before t_i and it obeys purely morphological principles.

If Hypothesis 2 holds, a completely random distribution of the synchronic types has to be expected, and it would be hopeless to look for diachronic explanations of the synchronic mismatches. Adopting Hawkins' viewpoint sketched above, we might say that Hypothesis 2 denies the Double Acquisition Hypothesis whereby if the constituent order in compounds changes then it must not do so in the absolute absence of the corresponding syntactic configuration. In other words, Hypothesis 2 would admit a diachronic scenario like the one in (1c) above.

However, while Hypothesis 1 is indeed testable, Harris and Campbell (1995, p. 210) are perfectly right when they conclude:

No one has yet presented evidence sufficient to indicate that the order of elements in a compound necessarily reflects the order of separate words in phrases at an earlier period or indeed to establish what relationship(s) there is (are) between the order of words in a phrase at a given time and the order in compounds. Until a systematic study is carried out ..., no conclusion can be reached about this relationship.

The aim of this paper is to present empirical evidence sufficient to solve the question. In order to assess which one of the two hypotheses is empirically supported, a systematic study of a sample of 50 languages has been carried out, in which the relation between the linearization of the constituents in syntax and in compounding has been investigated.

⁵ Or even worse: as an anonymous reviewer points out, it might also be that in a language a synchronic correspondence of constituent order in compounds and syntax be the result of an intermediate unattested stage of mismatch, subsequently eliminated. This can never be excluded and calls into question the reliability of the whole enterprise. However, the minimal (let's say "Galilean") assumption that I will give for presupposed here is that, given a system L observed during a certain temporal span, if a certain structure S_L occurs at both stages t_1 and t_2 , then in the absence of any proof testifying of the contrary nothing is likely to have happened to S_L .

What might be the worthiness of proving Hypothesis 1, if its reconstructive capacity remains weak, or even useless? Besides the value of its empirical content, proving this hypothesis may help us shedding light on the question of the autonomy of morphology as discussed in Sect. 2. If Hypothesis 1 turned out to be true, the emphasis on the superiority of morphological principles governing the constituent order in compounds would reveal its flimsiness. The following investigation will center on this aspect. Accordingly, putative mismatches falsifying Hypothesis 1 would be those cases of constituent order in compounds which could not be explained by making reference to (earlier) syntactic order, but should be motivated by inner morphological principles.

The latter are usually invoked in morphology when reordering within affix chains is observed. The best-known case is the so-called externalization of inflection, whereby an inflectional affix tends to be reordered when it happens to be ‘trapped’ by the grammaticalization of another morpheme in a dispreferred position like in Old Latin *e-um-pse* ‘he-ACC-self’ > Classical Latin *ips-um* ‘himself’. Such reorderings are easily explained by making reference to universal preferences of affix linearization with respect to the lexical stem (cf. Harris and Faarlund 2006), and/or to general principles like Bybee’s (1985) relevance. If Hypothesis 1 were falsified by the empirical findings, we should invoke similar linearization principles of morphological nature for explaining the constituent order in compounds, as suggested by the morphological literature mentioned in Sect. 2 above. Therefore, the investigation to be presented below aims at verifying both parts of the hypotheses, namely (i) the chaotic distribution of constituent order in compounds and syntax, and (ii) the feasibility of a syntactic or morphological explanation of the occurring mismatches.

4 Methodological aspects of sampling compounds

4.1 Problems with defining compounds

Defining compounds is not an easy task (cf. Bauer 2001; Aikhenvald 2007, pp. 24–35 for critical surveys). Leaving other intricacies aside, the main problem consists in identifying compounds with respect to other, phrase-like, complex structures.⁶ Generally, one relies on either formal or semantic criteria. Formal criteria are more robust, because they identify substantial properties which oppose compounds to phrases, as for instance stress in the well-known English pair *bláck bòard* vs. *bláck bóard* (see however Giegerich 2004 for a careful investigation of the phenomenon). A further example comes from Maori (cf. Bauer 1993, p. 518) and refers to lexical integrity (cf. Bresnan and Mchombo 1995 for an insightful

⁶ On the other hand, it is also problematic to distinguish between compounding and derivation. In this regard, Bauer (2005, p. 97) concludes that “the borderline between the two types is nevertheless as well established as the notion of word is”. In virtue of the difficulty of defining a word as such, this conclusion is paradoxically quite comforting: like in every discipline, basic notions in linguistics are not less helpful because of their conceptual density.

discussion). In Maori, adjective coordination in a noun phrase implies the repetition of the nominal or of a similar head:

- (3) a. *teetahi tangata tino nui, tangata tino moomona*
 a(sp) man very big man very fat
 ‘a big, fat man’
 b. *he koowhatu nui, he mea taimaha*
 a stone big a thing heavy
 ‘a big, heavy stone’

Accordingly, the following example must be considered a compound, because the repetition of the head is not allowed and the lexical integrity of the unit is preserved:

- (4) a. *te whare heihei nui* b. **te whare heihei, te mea nui*
 the house hen big the house hen the thing big
 ‘the large hen-house’

Far less robust is the semantic criterion, which identifies compounds on the basis of the non-compositionality of meaning, as again testified by the English compound *bláck bòard*. How far the meaning of a complex structure has to be considered idiomatized, is however highly debatable, because, to mention only one reason, one has to strongly rely on the subjective awareness of the individual researcher. Thus, independent morphological criteria should be sought for discriminating between true compounds and lexicalized phrases or listemes, which can only be made on a language-specific basis (cf. Gaeta and Ricca in press for a detailed discussion of the question).⁷ At any rate, the following two cases of increasing idiomatization can be mentioned, which show the kind of data that can be gathered from grammars, respectively of Tahitian (Lazard and Peltzer 2000) and of Amele (Roberts 1987, p. 330):

- (5) a. *pape to’eto’e* ‘ice’ b. *dodol gee-g* ‘larynx’
 water cold throat penis

Grammars are the main source of the research presented here. This choice may be questionable, because grammars often do not provide precise information on the specific properties of compounds, and simply list a number of examples collected in a purely impressionistic way. For instance, it has not seldom been difficult to get insights into the specific properties of compounds (allomorphies involved, productivity of the single patterns, etc.). Nonetheless, this choice presents the advan-

⁷ Moreover, the issue of productivity should also been seriously taken into consideration, as observed by an anonymous reviewer. True, my approach in this research has been quite liberal, including into my sample what the grammars I consulted reasonably considered to be compounds. Although such objections of a methodological nature are effective, I don’t think that they seriously undermine the specific perspective endorsed here. Even if some of the compounds included into the sample were likely to be excluded after a more accurate analysis, the reverse will never be true. Namely, that my “liberal” approach overlooks some true compounds which might potentially provide substantive evidence for the hypotheses laid down above. Thus, whatever the result of my investigation is, no potentially significant data are omitted.

tage of covering a broad spectrum of languages for which it would not be easy to have access to native speakers and to directly collect sufficient informations. Furthermore, for the specific purposes of this research, namely to assess the constituent order in compounds and phrases, important categorial properties of compounds like for instance endocentricity were not immediately relevant. Even an exocentric compound like *redskin* does contain the information of interest here, namely the adjective-noun order of the constituents. In this sense, I will speak of a head intended as a modified or governing element within a compound, regardless of the head role played in the compound as a whole.

A further important methodological point consists in distinguishing between the different types of constituents involved in the compounds. At least within theoretical morphology, it is often claimed that compounds are either right- or left-headed.⁸ This is misleading, because there are clearly different constituent types involved, which need not behave uniformly. In typological literature, this goes under the label of harmony (cf. Croft 2003, p. 62). In this research, the different compound types have been investigated separately, in order to assess their relation with the corresponding syntactic types. Following these guidelines, I investigated the constituent order within the NP, in particular between nominal heads and respectively adjectival and genitival modifiers, as well as the constituent order in the simple sentence, in particular between verb and objects.⁹ The constituent order of the different compound types have been checked against the corresponding syntactic types also elicited from grammars.¹⁰

It is important to keep this methodological aspect in mind, because the syntactic and the morphological patterns may be intrinsically deviant in some respect. For instance, the syntactic type involving subject (or external argument) and verb was not

⁸ For instance, cf. Fabb (1998, p. 67): “endocentric compounds tend to have heads in a language systematically on either the right (e.g. English) or left (e.g., Vietnamese, French)”. Again, we are not told what the empirical base of this claim is.

⁹ Since I only considered those compounds which categorially belong to nouns (either endocentric like *taxi driver* or exocentric like *pickpocket*), I completely excluded from my investigation noun incorporation defined as a morphological structure “in which a nominal constituent is added to a verbal root, and the resulting construction is both a verb and a single word” (Aikhenvald 2007, p. 11). I am aware that this choice is far from being unproblematic also because of cases like *talent-spotter* or *proof-reading*, which might be derived on the basis of an intermediate incorporation stage provided by the verbs *to talent-spot* and *to proof-read*. However, at least in English these incorporating verbs can be explained away as back-formations from the noun-noun compounds, and in this light they provide a good piece evidence for assessing the productivity of this compound type (cf. Plag 2003, p. 154, and Wurzel 1998 on similar types of incorporation in German, in particular back-formations like *Bergsteiger* ‘mountain-climber’ → *bergsteigen* ‘lit. to mountain-climb’). Furthermore, even if incorporation turned out to matter, because some cases like *proof-reading* were likely to be derived from a verb *to proof-read*, the effect of this on my data collection would be simply that the number of compounds would be reduced. Again, no potentially relevant data would be omitted. Clearly, the role of noun incorporation remains to be better understood, but this has to be left open for future research.

¹⁰ Clearly, there may be problems as for the reliability of these data similar to what has been discussed above for compounds. Moreover, in some cases languages display different and competing word orders, and it is not easy to decide what has to be considered basic (cf. Dryer 2007 for a critical discussion). This is especially true for the verb/object order. However, I will follow the practice established in typological research and I will trust the descriptive grammars of the single languages available to me, being confident that minor inaccuracies will not distort the global picture.

taken into consideration, because it contains an intrinsic ambiguity as for its compound correspondent. In SVO languages like Wakerena (Aikhenvald 1998) or Khmer (Antelme 2004) we find compounds of the type [N+V] as shown respectively below:

- (6) a. *mawaʎi-ʃia* ‘snake den’ b. *dɛj saa* ‘chalk’
 snake-inhabit earth be:white

Clearly, this pattern is at odds with the general [V+N] compound type occurring in these languages and rather corresponds to the [N+V] compound type generally occurring in head-final languages. In other words, considering the external argument introduces a further variation parameter, which is far from being clarified by the typological literature, namely the status of the SVO type with respect to the other harmonic types VSO and SOV (cf. Croft 2003, p. 72). For these reasons, this syntactic type and its compound companion have not been taken into consideration for this research.

4.2 The language sample

The sample on which this research is based includes fifty languages, which belong to a wide number of families, as can be gathered from Table 1.

Table 1 The language sample

Afro-Asiatic (5) Amharic, Modern Standard Arabic, Ge’ez, Hausa, Berber (Tamazight)	10%
Algic (1) Eastern Cree	2%
Altaic (2) Khalkha Mongolian, Turkish	4%
Arawakan (1) Warekena	2%
Australian (2) Djapu, Pitta Pitta	4%
Austro-Asiatic (1) Khmer	2%
Austronesian (8) Cèmuhí, Maori, Nêlêmwa, Palauan, Rapanui, Samoan, Tagalog, Tahitian	16%
Basque (1) Basque	2%
Burushaski (1) Burushaski	2%
Dravidian (1) Classical Tamil	2%
Eskimo-Aleut (1) West Greenlandic	2%
Indo-European (5) Armenian, Modern Greek, Kurdish, Persian, Welsh	10%
Kartvelian (1) Georgian	2%
Korean (1) Korean	2%
Mura (1) Pirahã	2%
Nakh-Daghestanian (2) Ingush, Udi	4%
Niger-Congo (6) Bambara, Ewe, Mbili, Swahili, Tswana, Yoruba	12%
Nilo-Saharan (2) Kanuri, Turkana	4%
Nivkh (1) Nivkh	2%
North Caucasian (1) Abkhaz	2%
Quechuan (1) Quechua	2%
Trans-New Guinea (1) Amele	2%
Uralic (4) Estonian, Finnish, Hungarian, Udmurt	8%

In collecting the sample, I followed Koptjevskaja-Tamm's (1993, p. 82) idea of a convenience sample, which is particularly useful when it is difficult to determine the object of the study and the available information is rather uneven. Besides such methodological reasons, the choice of collecting a sample similar to Koptjevskaja-Tamm's was also motivated by the idea of having results comparable to those obtained for nominalizations, namely for another significant morphological process, to be exploited for future research. Moreover, data for a number of languages were not accessible to me, so that I had to partially change the sample by excluding some languages investigated by Koptjevskaja-Tamm. In fact, her sample was larger, because it contained some 70 languages. However, I tried to preserve the same wide coverage of language families, even though I decided to discard some more Indo-European languages included in Koptjevskaja-Tamm's sample. This was due to the diachronic orientation of my research, in that I wanted to maximize the distribution across the language families, in order to exclude possible diachronic bias due to specific language families. A possible drawback of this sample comes from the distribution in the six macroareas which is rather uneven (Table 2).

Even if the distribution across the six macroareas is not optimal in that Eurasia and Africa are overrepresented compared to the others, I think that the wide number of families represented allows me to aspire to a certain degree of reliability as for the results obtained.

Furthermore, the sample is quite balanced with respect to the dominant word order types attested at least for two of the investigated syntactic types (Table 3). In this light, even if areality is not completely kept under control as a variable, the sample is pretty well balanced for two syntactic types investigated. Therefore, both

Table 2 Macroareas represented in the sample

Macroareas	
Africa	13
Australia-New Guinea	3
Eurasia	20
North America	2
SE Asia & Oceania	9

Table 3 Word order types attested in the sample

Word order type	Lgs.
OV	23
VO	25
varied	2
[Gen N] _{NP}	26
[N Gen] _{NP}	23
varied	1
[Adj N] _{NP}	19
[N Adj] _{NP}	29
varied/irrelevant	2

5 Results: constituent order in compounds and syntax

In table 4 and 5, the rows printed in bold give the harmonic types, whereas the first and the fourth lines contain the mismatches, namely the types violating the compound/syntax harmony. As can be gathered from the tables, the deviant cases are by far outnumbered by the harmonic types, which represent respectively 80% and 90% of the languages. Nevertheless, they exist. Let us see whether they may be treated as cases explainable in terms of diachronic drift.

$[N+N]_N$ & $[N \text{ Gen}]_{NP}$	4/8%
$[N+\bar{N}]_N$ & $[N \text{ Gen}]_{NP}$	20/42%
$[\bar{N}+N]_N$ & $[\text{Gen } N]_{NP}$	23/48%
$[N+\bar{N}]_N$ & $[\text{Gen } N]_{NP}$	1/2%
Total	48/100%

The first deviant case is represented by those languages, in which the syntactic type is left-headed, whereas the compound is right-headed:

- (8) a. [Adj+N]_N & [N Adj]_{NP} Basque, Mbili, Persian, Welsh
 b. [N+N] & [N Gen]_{NP} Modern Greek, Kanuri, Persian, Welsh

Let us focus on the first deviant type, in which we find [Adj+N]_N & [N Adj]_{NP}. As for Basque, we observe (cf. Coyos 2004) that the normal compound type is left-headed, as shown in (9a):

- (9) a. *egit.andi* 'exploit'
 do.great
 b. *euskal.tzain* 'Academician of the Basque language'
 Basque.guardian

The type given in (9b), which deviates from the syntactic order [N Adj]_{NP}, is mainly constituted by a sort of "préfixes à contenu locatif plutôt que des adjectifs autonomes" (Coyos 2004, p. 63), like *euskal*- 'Basque' which has to be related to *euskara*, or *erdal* in *erdal gramatika* 'foreign grammar' from *erdara* 'foreigner'.

In Mbili (cf. Ayuninjam 1998, p. 210), on the other hand, both constituent orders are attested for compounds:

- (10) a. *lwan ɲwɛn* 'old man'
 old person
 b. *məbɛ ndɪŋkaŋ* 'grandmother'
 mother elderly

Although the dominant syntactic type is head-initial, the mirror-image syntactic type [Adj N]_{NP} also occurs in a limited way. In Persian (cf. Lazard 2006), both linearizations occur with the compounds:

- (11) a. *siâh-češm* 'black-eyed'
 black-eye
 b. *leng-derâz* 'long-legged'
 leg-long

The dominant type is the more recent left-headed *ezâfé* construction [N-*e* Adj]_{NP}, which is however accompanied by the rarer and older right-headed type:

- (12) a. *pedar-e bičâre-at* 'your poor father'
 father-e poor-your
 b. *bičâre pedar-at*
 poor father-your

Finally, in Welsh (Awbery 2004), in which the normal syntactic type is [N Adj]_{NP}, both left- and right-headed adjective noun compounds are attested:

- (13) a. *gair mwys* 'pun' b. *bras-lun* 'outline'
 word ambiguous rough-picture

Also in Welsh the disharmonic type can be taken to be a remnant of an older syntactic type, replaced by the matching left-headed construct. In this regard, Awbery (2004, p. 307) observes that “non-phrase-like compounds [namely the type in (13b)] are not a very productive category in Modern Welsh”. In other words, the lexical drift has the effect of preserving non-productive word formation types, which are predicted to slowly fade away (cf. Gaeta 2008).

As for the other disharmonic type $[N+N]$ & $[N \text{ Gen}]_{NP}$, we find Kanuri and Modern Greek besides Persian and Welsh, for which the same analysis of Adj-N compounds can be repeated. In Kanuri (Cyffer 1998), the normal compound is right-headed, while the genitive construction is left-headed:

- (14) a. *súro-zàu* ‘stomach-ache’ b. *tátà kámu-vè* ‘woman’s son’
 stomach-pain son woman-GEN

However, Cyffer (1998, p. 50) adds in a footnote that: “[o]ccasionally one can observe that the genitive noun phrase precedes the head noun”, as in *Músa-be táda-nzé* ‘Musa’s son’. Thus, one may lay down the hypothesis that the $[N+N]$ order presumably mirrors an older syntactic type $[\text{Gen } N]_{NP}$ still preserved as a possible option for inalienable possession, as testified by similar changes observed in languages for which empirical evidence is available (e.g., in Maltese, cf. Koptjevskaja-Tamm 1996). It is clear that in the absence of diachronic data for Kanuri, this working hypothesis is of a high theoretical and empirical value. Notice that the “aberrant typological character of the Kanuri noun phrase” with respect to the universal harmonic tendencies is emphasized by Hutchison (1981, p. 8), because Kanuri also is strictly verb-final. In this light, a diachronic analysis of Kanuri appears even more mandatory.¹¹

In Modern Greek (cf. Holton et al. 1997, pp. 184, 264), the head-final type $[N+N]_N$ cooccurs with a dominant head-initial $[N \text{ Gen}]_{NP}$ and a more restricted head-final $[\text{Gen } N]_{NP}$ type:¹²

- (15) a. *to spíti tis Lukías* ‘Lucy’s house’
 the house the:GEN Lucy:GEN’
 b. *tis Lukías to spíti* ‘Lucy’s house
 [rather than anyone else’s]’
 c. *kapno-viomixanía* ‘tobacco industry’
 tobacco-industry

As for the other mirror-image disharmonic types, a similar picture emerges:

¹¹ See in this regard Cyffer (1997, p. 18): “Another source of comparable age is the Kanembu annotations of the Koran ... It is estimated that the annotations, called ‘*tafsir*’, are dated to the middle of the seventeenth century. However, no linguistic analysis has been made so far”.

¹² Cf. Holton et al. (1997, p. 264): “When the item in the genitive is a noun phrase (and especially when the head is definite), this noun phrase in the genitive may precede the head, generally for purposes of contrastive emphasis”. It should be added that the Modern (and Ancient) Greek compounds also match the constituent order reconstructed for the Indo-European compounds, which in their turn were harmonic with the reconstructed genitive/noun order (cf. Lehmann 1974, pp. 75–82, and Lindner 2002 for a survey). In this light, in Modern Greek, on a par with English, compounds may also be seen as a case for what will be called lexical conservatism in the next section.

- (16) a. [N+Adj]_N & [Adj N]_{NP}: Amharic, Armenian, Georgian, Udi
 b. [N+N] & [Gen N]_{NP}: Amharic

Amharic (cf. Leslau 1995) is interesting because it allows one to reconstruct the nest of linearization changes which took place in the course of the diachronic development from the Ge'ez mother tongue. Accordingly, the harmonic type corresponding to the right-headed syntactic types competes with the older compound types, mirroring the older left-headed construct states, typical of Semitic languages, testified by the linking morpheme *-ä*:

- (17) a. *posta bet* 'post office' b. *bet-ä krəstiyan* 'church'
 post house house-CN Christian

The normal genitival construction in Amharic is however right-headed:

- (18) *yä-bet-u mäskot täsäbr^wall* 'the window of the house is broken'
 of-house-the window is:broken

Also the other type at stake here is a remnant of the older construct state:

- (19) *angät-ä räğğim* 'long-necked'
 neck-CN long

The lexical drift is again testified by Armenian (cf. Donabédian 2004, p. 15), in which the disharmonic type is lexically restricted with respect to the right-headed harmonic type, in that the former occurs with cran-morphs like *-geγ* in (20b):

- (20) a. *sev-a-mort'* 'black person' b. *anjn-a-geγ* 'beautiful person'
 black-LK-skin person-LK-beautiful

In Georgian (cf. Hewitt 1995, p. 113, Harris and Campbell 1995, pp. 202–203), the word order changed from [N Adj]_{NP} to [Adj N]_{NP}, which explains the two possible types attested:

- (21) a. *did-gvar+i* 'aristocrat'
 big-surname
 b. *gul-ğ+ci+a* 'frank'
 heart-open

Table 6 VPs and compounds

[N+V] & VO	5/12%
[V+N] & VO	15/36%
[N+V] & OV	22/52%
[V+N] & OV	–
Total	42/100%

(22) a. *kala-gergec* 'cathedral' b. *ağbat-xeir* 'peace, greetings'
great-church fate-good

The last type which has been investigated is also in accordance with what has been discussed so far. As shown by Table 6, the large majority (88%) of the languages investigated display a harmonic order. Furthermore, one possible deviant case is not attested, namely [V+N] & *OV, and the other deviant type is represented by the following languages:

Ewe is characterized by a mixed word order, in that SOV occurs in clauses containing non-past tenses, while SVO is shown by clauses with past tenses (cf. Pasch 1995). Estonian and Finnish display a rather free word order, in which verb-final clauses are frequent, but the other types are also attested. Similar observations also hold for Modern Greek. Finally, for Welsh the same can be repeated that was observed above for noun-headed compounds.

It is undeniable that Hypothesis 2 is disconfirmed by the results. A strict relation occurs between word order types and head-modifier position in compounds, which strongly supports Hypothesis 1. The deviant cases lend support for an interpretation in terms of diachronic relations between language stages displaying different structural properties. Two paramount cases of today's frogs but yesterday's princes are Amharic and Kanuri. For Amharic, the synchronic disharmony may be fairly well reconducted to its Ge'ez mother tongue, while a similar pattern can only be hypothesized in the case of Kanuri, for which, however, diachronic evidence is lacking or still deserves a careful analysis.

(24) *[V+N] & OV

¹³ A possible counterexample is given by German (or Dutch) formations like *Habenichts* 'pauper, lit. have-nothing', which are however discarded as "not compounds at all" by Becker (1992, p. 27). Moreover, in German (and in Dutch) both linearization types occur, namely the VO type in main clauses and the OV types in embedded clauses. Again, the deviant compound type does not occur in the complete absence of any correspondent syntactic type.

logical consistency of (rigid) SOV languages (cf. Hawkins 1990, Gaeta and Luraghi 2001).

One important theoretical implication of the correspondence hypothesis is that it can be fairly well accommodated by a developmental theory like Hawkins' Double Acquisition Hypothesis, which forbids that a certain property P, whose occurrence implicates another property Q, is acquired before Q. Accordingly, an evolutionary stage is required in which either Q is already present at the earlier stage (for instance as a doublet of $-Q$) or P is acquired as doubling structure simultaneously with Q:

- (25) a. $-P \ \& \ -Q \ > \ *P \ \& \ -Q \ > \ P \ \& \ Q$
 b. $-P \ \& \ -Q \ > \ -P \ \& \ (-Q \sim) \ Q \ > \ P \ \& \ Q$
 c. $-P \ \& \ -Q \ > \ P \ \& \ Q$

In this vein, we may sketch two possible diachronic scenarios for the development of for instance the head-final type, and exclude a third one, in which the constituent order in compounds changes independently of the syntactic order (Table 7).

Table 7 Possible diachronic scenarios for the head-final type

$[N+N] \ \& \ [Gen \ N]_{NP}$

Kanuri

- > a. $[N+N] \ \& \ [Gen \ N]_{NP} \sim [N \ Gen]_{NP} >$
 $[N+N] \ \& \ ([Gen \ N]_{NP} \sim) [N \ Gen]_{NP}$

Welsh

- > b. $[N+N] \ \& \ [Gen \ N]_{NP} \sim [N \ Gen]_{NP} >$
 $[N+N] \sim [N+N] \ \& \ [Gen \ N]_{NP} \sim [N \ Gen]_{NP} >$
 $[N+N] \sim [N+N] \ \& \ [N \ Gen]_{NP}$
 > c. $*[N+N] \sim [N+N] \ \& \ [Gen \ N]_{NP} > [N+N] \sim [N+N] \ \& \ [N \ Gen]_{NP}$

The two scenarios are witnessed by Kanuri and Welsh, that display intermediate stages in which both $[Gen \ N]_{NP}$ and $[N \ Gen]_{NP}$ were available. Compared to Kanuri, Welsh developed a similar variation for the compounds, which is still preserved after the disappearance of the syntactic optionality. The mirror-image scenario is given by Amharic, in which, again, the head-initial constituent order changed first in syntax, and then parasitically in compounds (Table 8).

Table 8 Possible diachronic scenarios for the head-initial type

$[N+N] \ \& \ [N \ Gen]_{NP}$

Amharic

- > a. $[N+N] \ \& \ [N \ Gen]_{NP} \sim [Gen \ N]_{NP} >$
 $[N+N] \sim [N+N] \ \& \ [N \ Gen]_{NP} \sim [Gen \ N]_{NP} >$
 $[N+N] \sim [N+N] \ \& \ [Gen \ N]_{NP}$
 > b. $*[N+N] \sim [N+N] \ \& \ [N \ Gen]_{NP} > [N+N] \sim [N+N] \ \& \ [Gen \ N]_{NP}$

Admittedly, this hypothesis does not definitely solve Harris and Campbell's reconstruction problem: We still don't know how far back we have to go for reconstructing. However, we can relate a dyadic order reconstructed in this way to other syntactic features of a language. In other words, we have defined "an upper bound on language change" by constraining "the timing of changes relative to one another in interesting ways" (Hawkins 1983, p. 244).

A further important theoretical implication is that no independent constituent order change can be assumed for compounds. In this sense, morphology is not autonomous from syntax. In contrast to the slight preferences observed by Bauer for a constituent linearization in compounds which is independent from syntax and harmonic with the other compound types, the diachronic dimension gives us a different picture. Syntax seems to be the motor of the change, which may be then reflected in compounds.

This is confirmed by the analysis of the purported universal preference for Williams' Right-Hand Head Rule.

As can be gathered from the Table 9, the evidence is too scanty for assuming such a strong preference: with respect to the balanced number of head-initial and head-final possessive NPs, we only observe a small deviation in favor of the head-final type. This small deviation is represented by mismatches in which syntax gave the start to the innovation.

A further hint in support of the autonomy of compounds from syntax was the behavior of synthetic compounds, whose constituent order was observed by Bauer to be preferentially related to the *tatpuruṣa*, i.e. determinative compounds. Table 10 shows the data collected in my sample.

Table 9 Correlation between determinative compounds and NPs

[<u>N</u> +N]	21	[N Gen] _{NP}	23
[N+ <u>N</u>]	27	[Gen N] _{NP}	23
Compounds		Possessive NPs	

Table 10 Constituent order in synthetic and determinative compounds

[V+N] & [<u>N</u> + <u>N</u>]	1
[V + N] & [N + N]	14
[N + V] & [N + N]	22
[N+V] & [<u>N</u> + <u>N</u>]	3

Again, the harmonic types are given in bold, whereas the deviant cases are the following ones:

- (26) a. [V+N] & [N+N]: Welsh
 b. [N+V] & [N+N]: Amharic, Kurdish, Welsh

As expected, these can be explained in terms of diachronic scenarios, in which syntax is the motor of the change. We have already discussed the cases of Amharic and Welsh, but a similar explanation can also be provided for Kurdish, in which NPs

are consistently head-initial, whereas the basic word order in clauses can also be verb-final (cf. Bedir Khan and Lescot 1991).

The diachronic scenarios have to be seen in relation to other harmonic tendencies emerging from syntactic linearization. Accordingly, the constituent order of determinative compounds has to be connected with the harmonic preferences underlying the linearization of other syntactic types, and in particular the verb-object relation, which again underlies the synthetic compounds. In this light, the relation between determinative and synthetic compounds emerges as a more specific case within a broader generalization. And presumably, other preferences occur which may be explained in similar terms as well. For instance, we may consider the correlations among the three types investigated in this paper:

Table 11 Constituent order correlations among the three compound types

[V+N] & [N+N] & [N+A] _N	10	Berber, Khmer, Maori, Nêlêmwa, Samoan, Tagalog, Tahitian, Tswana, Welsh, Yoruba
[V+N] & [N+N] & [A+N] _N	2	Mbili, Welsh
[V+N] & [N+N] & [N+A] _N	1	Welsh
[V+N] & [N+N] & [A+N] _N	1	Welsh
[N+V] & [N+N] & [N+A] _N	3	Amharic, Kurdish, Welsh
[N+V] & [N+N] & [A+N] _N	1	Welsh
[N+V] & [N+N] & [N+A] _N	8	Abkhaz, Amele, Amharic, Armenian, Bambara, Basque, Pirahã, Welsh
[N+V] & [N+N] & [A+N] _N	12	Armenian, Basque, Classical Tamil, Estonian, Finnish, Hungarian, Modern Greek, Persian, Quechua, Turkish, Udmurt, Welsh

In Table 11 the harmonic correlations are given in bold: again, we observe a number of mismatches, which we have been already discussing so far. A further interesting finding of this table is the high number of deviant languages, in which the N-Adj compound differs from the other head-final types: [N+V] & [N+N] & [N+A]_N. Notice that this is expected on the basis of Hawkins' universal mentioned above in (2), in that the adjective/noun order is a bad predictor of harmonic linearization with respect to the genitive/noun order, whereas the reverse is true at least for the restriction *[Adj N]_{NP} & [N Gen]_{NP} in postpositional languages.¹⁴ In spite of

¹⁴ In spite of the general unreliability of the adjective/noun order with respect to cross-categorical order harmony, Dryer (1988, pp. 206–207) also finds Hawkins' correlation to be confirmed by his data. Furthermore, he mentions a counterexample to Hawkins' generalization, namely Djapu, which makes this generalization a statistical rather than an absolute universal. However, if I understand him correctly, Dryer does not discuss the potential problem constituted by this finding for his Branching Direction Theory (BDT), which should predict that no relevant correlations should occur involving the adjective/noun order. This problem also shows up in his (1992) paper, in which he observes that a more refined version of BDT is able to make "predictions not only about the ordering of heads with respect to dependents but also about the ordering of multiple dependents with respect to each other. If we assume, for example, that an N' consisting of an adjective, a noun, and a relative clause has a flat structure in the major constituent tree ..., then the [BDT, LG] predicts not only that the noun and the relative clause are a correlation pair, but also that the adjective and relative clause should be a correlation pair, because the former is a nonphrasal category and the latter is a fully recursive phrasal category" (Dryer 1992, p. 116). Given that the noun and the genitive also are a correlation pair, the same should apply here, therefore accounting for Hawkins' statistical universal. At any rate, I am rather sceptical about Dryer's assumption

the general unreliability of the adjective/noun order with respect to cross-categorical order harmony, Dryer (1988, pp. 206–207) also finds Hawkins' correlation to be confirmed by his data. Furthermore, he mentions a counterexample to Hawkins' generalization, namely Djapu, which makes this generalization a statistical rather than an absolute universal. However, if I understand him correctly, Dryer does not discuss the potential problem constituted by this finding for his Branching Direction Theory (BDT), which should predict that no relevant correlations should occur involving the adjective/noun order. This problem also shows up in his (1992) paper, in which he observes that a more refined version of BDT is able to make "predictions not only about the ordering of heads with respect to dependents but also about the ordering of multiple dependents with respect to each other. If we assume, for example, that an N' consisting of an adjective, a noun, and a relative clause has a flat structure in the major constituent tree ..., then the [BDT, LG] predicts not only that the noun and the relative clause are a correlation pair, but also that the adjective and relative clause should be a correlation pair, because the former is a nonphrasal category and the latter is a fully recursive phrasal category" (Dryer 1992, p. 116). Given that the noun and the genitive also are a correlation pair, the same should apply here, therefore accounting for Hawkins' statistical universal. At any rate, I am rather sceptical about Dryer's assumption that adjective phrases do not constitute a fully recursive phrasal category, because of the number of adjectives governing complements like *far*, *proper*, etc. Notice that in a consistent left-branching language like German, they are all placed at the left of the nominal head, as in these real examples from Google (query 14.2.2009): *oft ist es einfach ein mir fernes verhalten* 'often it is simply a behavior distant from me', *das ist ein mir eigenes ästhetisches Empfinden* 'this is an aesthetic feeling proper to me', which is at odds even with the refined version of BDT. This question needs a longer discussion and must be left open for future research. For our purposes, this implies that determinative compounds are likelier to reflect significant aspects of harmonic linearization than the others, and especially the adjective/noun compounds. In fact, the stock of languages displaying this deviation can be accounted for by referring to subsequent syntactic change (Amharic, Armenian and Welsh), and/or to a disharmony already present at the syntactic level (Abkhaz, Amele, Bambara, Basque and Pirahã). Again, all of this confirms our basic finding, namely that constituent order in compounds does not seem to change autonomously from syntax.

In this respect, compounds substantially differ from the rest of affixed words. While affixes may be reordered on the basis of strictly morphological principles as is arguably the case for 'trapped' affixes (cf. Harris and Faarlund 2006), lexical stems involved in compounds do not usually undergo any reordering, unless syntax has previously changed.

Footnote 14 continued

that adjective phrases do not constitute a fully recursive phrasal category, because of the number of adjectives governing complements like *far*, *proper*, etc. Notice that in a consistent left-branching language like German, they are all placed at the left of the nominal head, as in these real examples from Google (query 14.2.2009): *oft ist es einfach ein mir fernes verhalten* 'often it is simply a behavior distant from me', *das ist ein mir eigenes ästhetisches Empfinden* 'this is an aesthetic feeling proper to me', which is at odds even with the refined version of BDT. This question needs a longer discussion and must be left open for future research.

7 Why mismatches?

As a way of conclusion, we may finally ask why such mismatches arise at all. Given the violation of a synchronic constraint resulting from an implicational universal like ‘If P then Q’, they should never have come into existence. Why are they found? It should not be forgotten that the diachronic scenarios depicted above tell us that the mismatches are not the direct product of a change, but a sort of side-effect. Accordingly, they are inert, in the sense that their constituent order does not change independently of syntax. In other words, they are not acquired in a strict sense. To understand why mismatches are actually observed, one has to adopt a developmental perspective. A type which is excluded in a certain developmental scenario is a possible stage in its mirror-image counterpart. Let me be more explicit with the help of the example of determinative compounds and genitive/noun constructions:

- (27) a. $[N+N] \ \& \ [Gen \ N]_{NP} > [N+N] \ \& \ [N \ Gen]_{NP} > [N+N] \ \& \ [N \ Gen]_{NP}$
 b. $[N+N] \ \& \ [Gen \ N]_{NP} > *[\underline{N+N}] \ \& \ [\underline{Gen \ N}]_{NP} > [N+N] \ \& \ [N \ Gen]_{NP}$
 c. $[\underline{N+N}] \ \& \ [N \ Gen]_{NP} > [\underline{N+N}] \ \& \ [Gen \ N]_{NP} > [N+N] \ \& \ [Gen \ N]_{NP}$
 d. $[\underline{N+N}] \ \& \ [N \ Gen]_{NP} > *[\underline{N+N}] \ \& \ [\underline{N \ Gen}]_{NP} > [N+N] \ \& \ [Gen \ N]_{NP}$

The developmental stages in (27b) and (27d) marked with bold are not allowed. However, they correspond to transitional stages which are allowed by their respective mirror-image developmental scenarios in (27c) and (27a). In other words, the mismatches are not universally impossible. They are only developmentally impossible in a given diachronic scenario.

Nonetheless, their existence is clearly dysfunctional, in that they are disharmonic. We might ask why mismatches are preserved in spite of the syntactic development. A trivial answer is implicitly contained in Comrie’s quotation reported in Sect. 3 above, and calls into play the concept of lexical conservatism: since compounds are units entrenched in the lexicon, they are simply preserved, i.e. stored, as such in spite of their disharmonic effect. This does not necessarily imply that the lexicon is a store of irregularities, of listemes, as some scholars maintain (e.g., Di Sciullo and Williams 1987, p. 1). It may well be the case that the effects of a change in one component of the grammar are reflected in others in a slower way. For instance, it is fairly well known that the effects of phonological change are slowly ‘digested’ by morphology: analogical change is traditionally invoked as a morphological reaction against phonological opacity (cf. Gaeta 2007).

Can we go beyond this idea of lexical conservatism? One way to understand why such a dysfunctional state of affairs is preserved calls into play the divergent behavior of affixes which do undergo morphologically-driven reorderings. In this regard, it should be noted that lexical stems differ from affixes categorially, because they display inner referential strength. In this sense, they are different from grammatical objects like affixes, in that they cannot be organized into universal hierarchies obeying general principles like Bybee’s relevance. Moreover, such principles like relevance also contribute to modeling the inner design of morphological systems, which generally display a certain degree of paradigmatic consistency: accordingly, affixes may be systematically distributed in recurrent slot patterns,

which also possess an inner structural force. Clearly, lexical stems are much less sensitive to paradigmatic force. In the absence of such general motivations, strictly morphologically-driven reorderings are not expected to occur.

Finally, an interesting, more speculative, suggestion recently laid down by Jackendoff (2009) is that compounds are the relics of protolanguage. The latter is taken to be “a cognitive ‘scaffolding’ on which modern language is built” and “consists of ... a direct interface between phonology and semantics”, a sort of evolutionary step between unregulated concatenation and full syntax (cf. also Jackendoff 2002, pp. 245–251 for an extensive discussion). If one espouses this idea that modern language is “laid over” a protolinguistic substrate, “the intriguing possibility” arises “that the coverage is not complete: that there exist pockets of modern language that are relics of earlier stages of the language capacity”. Compounds may be such relics, namely “areas where there is only rudimentary grammatical structure, and in which such grammatical structure as there is does not so much to shape semantic interpretation”. This explains why compounds do “not look like the rest of morphology”: they are “actually not a *grammatical* phenomenon, but a *protogrammatical* one”.

In this light, compounds may be taken to tolerate deviant linearizations because they display “only rudimentary grammatical structure”, which as such “does not so much to shape semantic interpretation”. Furthermore, that compounds must be treated apart from the rest of morphology, or at least from affixal derivation, is also supported by our findings that speak against assuming inner-morphological linearization preferences for compounds in contrast with what seems to be the case of affixes. However, the picture cannot be so easy: we have indeed seen that the constituent order of compounds depends on syntax, and cannot simply be the result of “a direct interface of phonology and semantics” (recall Paul’s initial quote above). Rather, the mismatches might be claimed to represent pockets of modern language that are relics of earlier modeling of syntactic structure. Nonetheless, the suggestion of treating compounds as relics whose constituent order as such does not have any bear on the violation of synchronic constraints on grammars—in Jackendoff’s terms, compounds are “only barely syntactic”—can help us explain why the mismatches survive. Future research will tell us whether this suggestion is on the right track.

Acknowledgements The research presented in this paper was supported by the PRIN-project COMPONENT (2006–2007) coordinated by S. Scalise and funded by the Italian Ministry of University and Research (MUR). Parts of this research were presented at the 18th *International Conference on Historical Linguistics* (Montréal 6.–11.8.2007), and at the *Sixth Mediterranean Morphology Meeting* (MMM6, Ithaca 27.–30.9.2007). I am very grateful to the people present on these occasions, and especially Alice Harris and Sergio Scalise, as well as three anonymous reviewers and to the editor Ingo Plag, to whom I have accumulated an enormous debt of gratitude. Needless to say, opinions expressed and remaining mistakes are of my own responsibility.

References

- Aikhenvald, A. Y. (1998). Warekena. In D. C. Derbyshire, & G. K. Pullum (Eds.), *Handbook of Amazonian languages* (Vol. 4, pp. 225–439). Berlin: Mouton de Gruyter.

- Aikhenvald, A. Y. (2007). Typological distinctions in word-formation. In T. Shopen (Ed.), *Language typology and syntactic description* (Vol. 3: *Grammatical categories and the lexicon*, pp. 1–65.). 2nd ed. Cambridge: Cambridge University Press.
- Antelme, M. (2004). Khmer. In P. J. L. Arnaud (Ed.), *Le nom composé. Données sur seize langues* (pp. 149–183). Lyon: Presses Universitaires de Lyon.
- Awbery, G. M. (2004). Welsh. In P. J. L. Arnaud (Ed.), *Le nom composé. Données sur seize langues* (pp. 303–327). Lyon: Presses Universitaires de Lyon.
- Ayuninjam, F. F. (1998). *A reference grammar of Mbili*. Lanham: University Press of America.
- Bauer, W. (1993). *Maori*. London: Routledge.
- Bauer, L. (2001). Compounding. In M. Haspelmath, E. König, W. Österreicher, & W. Raible (Eds.), *Language typology and language universals* (pp. 695–707). Berlin & New York: Mouton de Gruyter.
- Bauer, L. (2005). The borderline between derivation and compounding. In W. U. Dressler, D. Kastovsky, O. E. Pfeiffer, & F. Rainer (Eds.), *Morphology and its demarcations. Selected papers from the 11th morphology meeting, Vienna, February 2004* (pp. 97–108). Amsterdam: John Benjamins.
- Becker, T. (1992). Compounding in German. *Rivista di Linguistica*, 4, 5–36.
- Bedir Khan, Dj., & Lescot, R. (1991). *Grammaire Kurde (Dialecte Kurmandji)*. Paris: Maisonneuve.
- Bresnan, J., & Mchombo, S. A. (1995). The lexical integrity principle: Evidence from Bantu. *Natural Language and Linguistic Theory*, 13, 181–254. doi:10.1007/BF00992782
- Bybee, J. (1985). *Morphology. The relation between meaning and sound*. Amsterdam: John Benjamins.
- Comrie, B. (1980). Morphology and word order reconstruction: Problems and prospects. In J. Fisiak (Ed.), *Historical morphology* (pp. 81–96). The Hague: Mouton.
- Coyos, J.-B. (2004). Basque. In P. J. L. Arnaud (Ed.), *Le nom composé. Données sur seize langues* (pp. 47–70). Lyon: Presses Universitaires de Lyon.
- Croft, W. (2003). *Typology and universals* (2nd ed.). Cambridge: Cambridge University Press.
- Cyffer, N. (1997). A survey of the Kanuri language. In N. Cyffer & T. Geider (Eds.), *Advances in Kanuri scholarship* (pp. 17–66). Köln: Köppe.
- Cyffer, N. (1998). *A sketch of Kanuri*. Köln: Köppe.
- Demske, U. (2001). *Merkmale und Relationen. Diachrone Studien zur Nominalphrase des Deutschen*. Berlin: Walter de Gruyter.
- Di Sciullo, A. M., & Williams, E. (1987). *On the definition of word*. Cambridge, MA: MIT Press.
- Donabédian, A. (2004). Arménien. In P. J. L. Arnaud (Ed.), *Le nom composé. Données sur seize langues* (pp. 3–20). Lyon: Presses Universitaires de Lyon.
- Dressler, W. U. (2005). Compound types. In G. Libben & G. Jarema (Eds.), *The representation and processing of compound nouns* (pp. 23–44). Oxford: Oxford University Press.
- Dryer, M. S. (1988). Object-verb order and adjective-noun order: Dispelling a myth. *Lingua*, 74, 185–217. doi:10.1016/0024-3841(88)90059-9
- Dryer, M. S. (1992). The Greenbergian word order correlations. *Language*, 68(1), 81–138.
- Dryer, M. S. (2007). Word order. In T. Shopen (Ed.), *Language typology and syntactic description* (Vol. 1: *Clause structure*, pp. 61–131.). 2nd ed. Cambridge: Cambridge University Press.
- Fabb, N. (1998). Compounding. In A. Spencer & A. M. Zwicky (Eds.), *The handbook of morphological theory* (pp. 66–83). Oxford: Blackwell.
- Gaeta, L. (2005). Word formation and typology: Which language universals? In G. Booij, E. Guevara, A. Ralli, S. Sgroi, & S. Scalise (Eds.), *Morphology and linguistic typology, on-line proceedings of the fourth mediterranean morphology meeting (MMM4) Catania 21–23 September 2003*, University of Bologna, 2005. Retrieved from <http://morbo.lingue.unibo.it/mmm/>
- Gaeta, L. (2007). Is analogy economic? In F. Montermini, G. Boyé, & N. Hathout (Eds.), *Selected Proceedings of the 5th Décembrettes: Morphology in Toulouse* (pp. 20–33). Somerville, MA: Cascadilla Press.
- Gaeta, L. (2008). Mismatch: grammar distortion and grammaticalization. In M. J. López-Couso & E. Seoane (Eds.), *Rethinking grammaticalization. New perspectives* (pp. 103–127). Amsterdam: John Benjamins.
- Gaeta, L., & Luraghi, S. (2001). Gapping in classical greek prose. *Studies in Language*, 25, 89–113.
- Gaeta, L., & Ricca, D. (in press). *Composita solvantur*: Compounds as lexical units or morphological objects? In L. Gaeta & M. Grossmann (Eds.), *Compounds between syntax and lexicon*. Special Issue of *Italian Journal of Linguistics/Rivista di Linguistica* 21.2.
- Giegerich, H. J. (2004). Compound or phrase? English noun-plus-noun constructions and the stress criterion. *English Language and Linguistics*, 8, 1–24.

- Harris, A. C., & Campbell, L. (1995). *Historical syntax in cross-linguistic perspective*. Cambridge: Cambridge University Press.
- Harris, A. C., & Faarlund, J. T. (2006). Trapped morphology. *Journal of Linguistics*, 42, 289–315.
- Harrison, S. P. (2003). On the limits of the comparative method. In B. D. Joseph & R. D. Janda (Eds.), *The handbook of historical linguistics* (pp. 213–243). Oxford: Blackwell.
- Hawkins, J. A. (1983). *Word order universals*. New York: Academic Press.
- Hawkins, J. A. (1990). Seeking motives for change in typological variation. In W. Croft, K. Denning, & S. Kemmer (Eds.), *Studies in typology and diachrony for Joseph H. Greenberg* (pp. 95–128). Amsterdam: John Benjamins.
- Hewitt, B. G. (1995). *Georgian. A structural reference grammar*. Amsterdam: John Benjamins.
- Holton, D., Mackridge, P., & Philippaki-Warbuton, I. (1997). *Greek. A comprehensive grammar of the modern language*. London: Routledge.
- Hutchison, J. P. (1981). *The Kanuri language. A reference grammar*. Madison: African Studies Program University of Madison.
- Jackendoff, R. (2002). *Foundations of language*. Oxford: Oxford University Press.
- Jackendoff, R. (2009). Compounding in the parallel architecture and conceptual semantics. In R. Lieber & P. Štekauer (Eds.), *The Oxford handbook of compounding*. Oxford: Oxford University Press.
- Koptjevskaja-Tamm, M. (1993). *Nominalizations*. London: Routledge.
- Koptjevskaja-Tamm, M. (1996). Possessive noun phrases in Maltese: Alienability, iconicity, and grammaticalization. *Rivista di Linguistica*, 8, 245–274.
- Lazard, G. (2006). *Grammaire du persan contemporain*. Paris: Institut français de recherche en Iran Persian.
- Lazard, G., & Peltzer, L. (2000). *Structure de la langue tahitienne*. Paris: Peeters.
- Lehmann, W. P. (1974). *Proto-Indo-European syntax*. Austin: University of Texas Press.
- Leslau, W. (1995). *Reference grammar of Amharic*. Wiesbaden: Harrassowitz.
- Lindner, T. (2002). Nominalkomposition und Syntax. In H. Hettrich (Ed.), *Indogermanische Syntax—Fragen und Perspektiven* (pp. 263–279). Wiesbaden: Reichert.
- Pasch, H. (1995). *Kurzgrammatik des Ewe*. Köln: Köppe.
- Paul, H. (1920/1995). *Prinzipien der Sprachgeschichte* (10th ed.). Tübingen: Niemeyer.
- Paul, H. (1920). *Deutsche Grammatik. Bd. 5: Wortbildungslehre*. Halle a.S.: Niemeyer.
- Plag, I. (2003). *Word-formation in English*. Cambridge: Cambridge University Press.
- Roberts, J. R. (1987). *Amele*. London: Croom Helm.
- Schulze, W. (2004). Udi. In P. J. L. Arnaud (Ed.), *Le nom composé. Données sur seize langues* (pp. 287–301). Lyon: Presses Universitaires de Lyon.
- Spencer, A. (1991). *Morphological theory*. Oxford: Blackwell.
- Williams, E. (1981). On the notions 'lexically-related' and 'head of a word'. *Linguistic Inquiry*, 12, 245–274.
- Wurzel, W. U. (1998). On the development of incorporating structures in German. In R. M. Hogg & L. van Bergen (Eds.), *Historical linguistics 1995* (Vol. 2: *Germanic Linguistics*, pp. 331–344). Amsterdam: John Benjamins.